What is data journalism?

- **Data journalism** reflects the increased role of numerical data for reporting in the digital era.
  - Reflects increased interaction between journalists and visualization design, computer science and statistics.
- Aspects:
  - Database analysis: FOI requests on topics like crime, racial discrimination, judicial sentencing, toxic pollution, ...
  - Visualization: infographic data vis., emphasis on the story
  - Increasing use & development of data vis. software tools
    - analysis: R, python, ...
    - visualization libraries: R, D3, javascript, ...
  - Increasing use of dynamic updating of data sources
    - election polls / prediction
    - disease tracking

Bill Dedman: *The Color of Money*

Bill Dedman of *The Atlanta Journal-Constitution* analyzed racial discrimination by banks and other lenders in middle-income black neighborhoods of Atlanta, GA.

In a series of articles, May 1-4, 1988, *The Color of Money*, he showed that although banks made loans in even the poorest white neighborhoods, they did not lend in middle-class or more affluent black neighborhoods.

This work won the Pulitzer prize in investigative reporting in 1989.

It helped establish the idea of computer-assisted reporting and the importance of compelling visualizations in telling the story.

See: [http://powerreporting.com/color/](http://powerreporting.com/color/) for all the articles and images

Hurricane Andrew

- In Aug. 1992, a category 5 hurricane hit Miami-Dade county
- Huge numbers of houses were destroyed, but this varied by neighborhood
- Investigative reporters at the *Miami Herald* discovered why:
  - Building codes and inspection became lax after 1980.
  - Older homes survived.
- This report won the Pulitzer Prize in 1993.

Source: [https://www.flickr.com/photos/juggernautco/sets/72157607210036175/](https://www.flickr.com/photos/juggernautco/sets/72157607210036175/)
Hurricane Andrew

- The report in the *Miami Herald* contained many detailed analysis maps.
- There was compelling evidence for the cause of destruction.
- The combination of investigative reporting, analysis and graphics was seen as powerful.

Predicting elections

- In 1962, John Tukey and a statistical team joined with NBC to attempt to predict the congressional election.
- Over time, they developed new methods to combine past results with exit polls and current returns to give up-to-date estimates and measures of uncertainty as the numbers came in.
- Key idea: “borrowing strength”
- Now known as “hierarchical Bayes”

CNN Magic Wall

Direct, interactive visualization of election news on TV took a big jump with the development of the Magic Wall, a large touch-screen display, tied to databases of election results, and with on-screen controls to zoom in/out, compare, highlight, etc.

Election reporting: NYT

In the special election for the Alabama U.S. Senate, Dec. 12, 2017, *The New York Times* used dynamic, interactive maps to compare the vote to the 2016 Presidential election.

- Hovering the mouse over a county showed the details.
- A map-based display is natural & easy.
- Q: What other graphic might show the comparison better?
Election reporting: Dashboards

Throughout the evening of Dec. 12, the NYT dynamically updated dashboard displays of the election returns and model-based predictions of the winner.

Another dynamically updated chart showed the projected vote margin, together with two levels (50%, 95%) of uncertainty intervals for the current estimate.

Novel graphic designs

This graph, from fivethirtyeight.com was designed to show how some presidential candidates had shifted positions before the 2016 election.

The axes are a score on social and economic policy, but they rotated the axes by 45° to create zones related to political thought.

This info graphic is eye-catching and self-explanatory:
- colored/labeled zones
- interpretive labels on axes
- arrows showing movement to extremes

Measles and vaccines

Visualizing the impact of health policy interventions

In 2015 Tynan DeBold & Dov Friedman in the Wall Street Journal tried to show the effect of the introduction of vaccination programs in the US states on disease incidence, using color-coded heat maps for a variety of diseases.

The long time series ~70 years made this work.

The heat map color scale is not exemplary, but the message is still clear: disease incidence declined after vaccines were introduced.

The images are “interactive,” in the weak sense that tool-tips are shown on mouse movement.

Making the message more explicit

What you should try to show here in an info graphic is the overall impact of vaccination on measles. Ed Tufte did this by adding a histogram at the top showing total # of cases by year.

With the overall impact made clear, details about individual states can now be explored.

Effect ordering & color scales

Ordering the states by an overall measure of measles incidence makes it easier to study details. A uni-dimensional scale of color is more effective than the ad-hoc scale used in the original.

What’s wrong in Wis. & VT?

This chart, by Xan Gregg also shows missing data explicitly.

Color scales for data graphs is a growing topic.

Best examples of DDJ? Ask Twitter

538: Atlas of Redistricting

There’s a lot of complaining about gerrymandering, but what should US Congressional districts look like? What would they look like to serve various goals?

This map shows the current boundaries, with expected wins by Rep and Dem.

Why do I find this app appealing?
How could district boundaries be re-drawn to promote more competitive elections?

Election district boundaries are consequential & highly political. They can be changed by state legislatures.

Different scenarios show the changes in representation, w/o a single voter switching parties!

Note that the overall impression of the map is still Rep. overall.

Changes in representation take place largely on the east/west coasts

Aaron Williams & Armand Emamdjomeh at the Washington Post studied changes in racial composition and diversity in U.S. cities over 30+ years. They relate these to demographic, policy, and urban/suburban issues using well-designed thematic maps.

Daniel Dale & Ed Tubb at the Toronto Star have maintained a database of every false claim made by Donald Trump since the Inauguration on Jan. 20, 2017.

Donald Trump has said 4557 false things as U.S. president (as of 3-13-2019)
Every false claim is documented: place, claim, truth, “slug”, categories

Graphic artist Jelle Geertsma designed this animated graphic showing the top 5 claim “slugs” & the top 20 categories over time, by month.

Bars at right show cumulative totals

This animation uses the gganimate package.

A TrumpLies package is in progress.